

Технологии Semantic Web

Семантика ресурсов: RDF, RDF Schema, RDFa

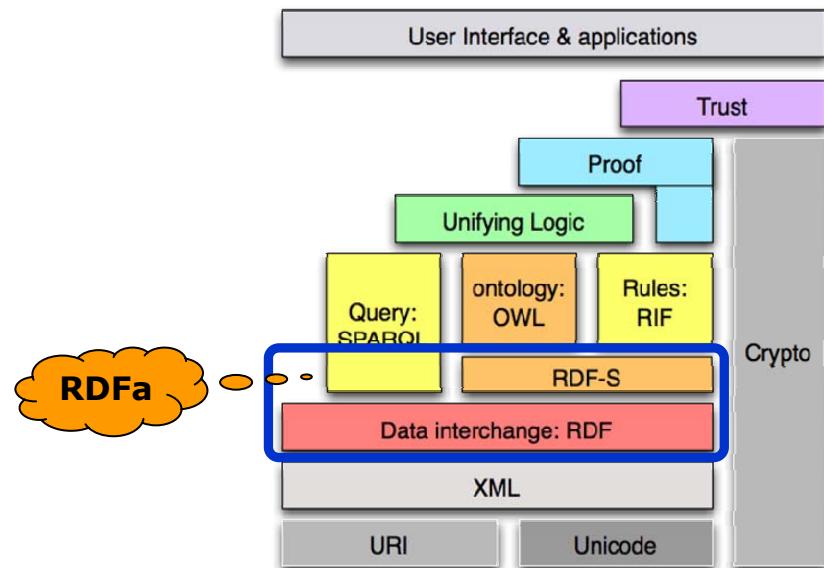
Ермоляев В. А.

Кафедра ИТ
Запорожский нац.
университет
Украина

ТЗ

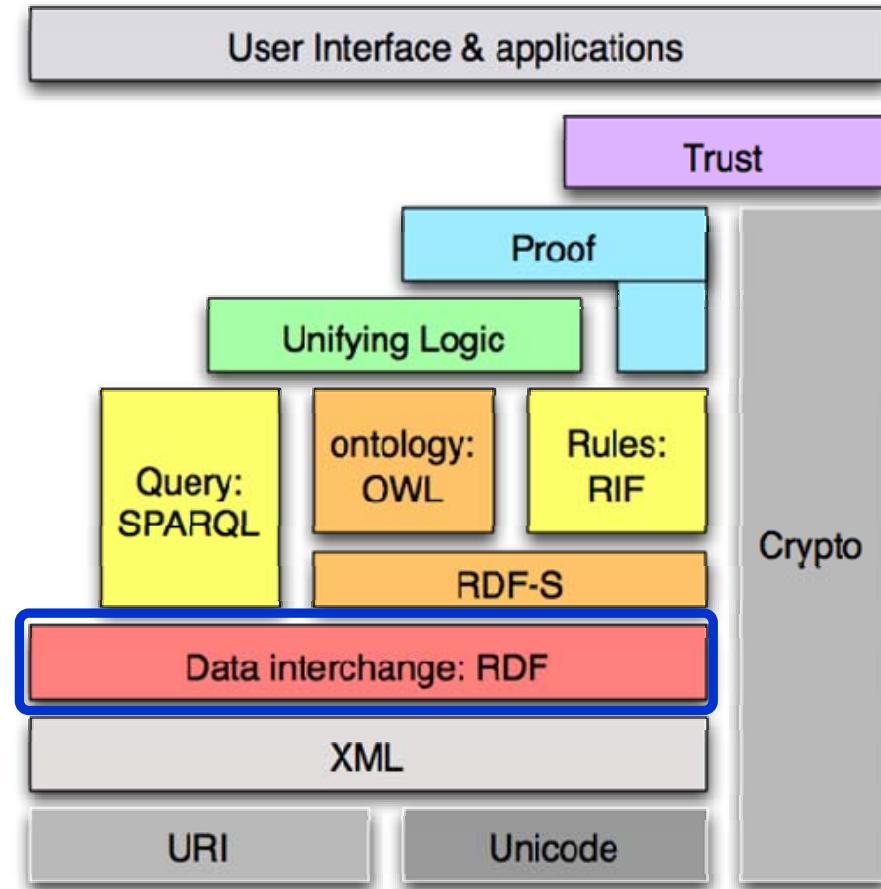


<http://ermolayev.com/>



План

- Обзор возможностей RDF
- Критика RDF
- XML синтаксис RDF



RDF и RDFS

□ **RDF:**

- Стандарт W3C специфицирующий нотацию для описания семантики Web ресурсов

□ **RDF Схема:**

- Стандарт W3C специфицирующий словарь (тезаурус - пространство терминов) для последующего использования в описаниях на RDF
- Организует тезаурус в виде типизированной иерархии
- Дает возможность однозначно специфицировать семантические отношения между терминами в тезаурусе

RDF – Первый Язык SW

object

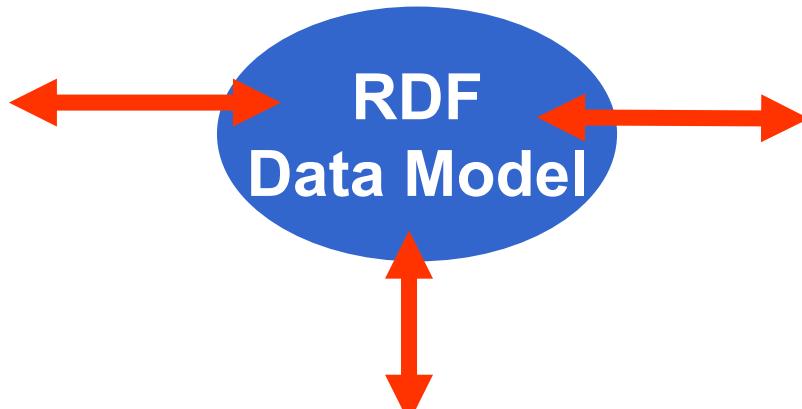
predicate

subject

Запись в XML

```
<rdf:RDF .....>
  <....>
  <....>
</rdf:RDF>
```

Удобна
для машинной
обработки



```
stmt(docInst, rdf_type, Document)
stmt(personInst, rdf_type, Person)
stmt(inroomInst, rdf_type, InRoom)
stmt(personInst, holding, docInst)
stmt(inroomInst, person, personInst)
```

Удобны для лог. вывода

Технологии Semantic Web, 2012

Граф

Удобен для
чедовека

*RDF - простой язык
для построения
представлений
(семантики) на базе
графов*

Недостатки XML

- XML – универсальный (мета-)язык для спецификации разметки
- XML предоставляет унифицированный механизм для обмена данными и мета-данными между приложениями
- Однако, XML **НЕ предоставляет** средств для спецификации семантики данных
- Пр.: С тегами (в т.ч. вложенными) не связывается никакого смысла
 - Разные приложения в полном праве интерпретировать таги (и вложенность) ПО-РАЗНОМУ

Вложенность тегов в XML

Victor Yanukovich is the president of Ukraine

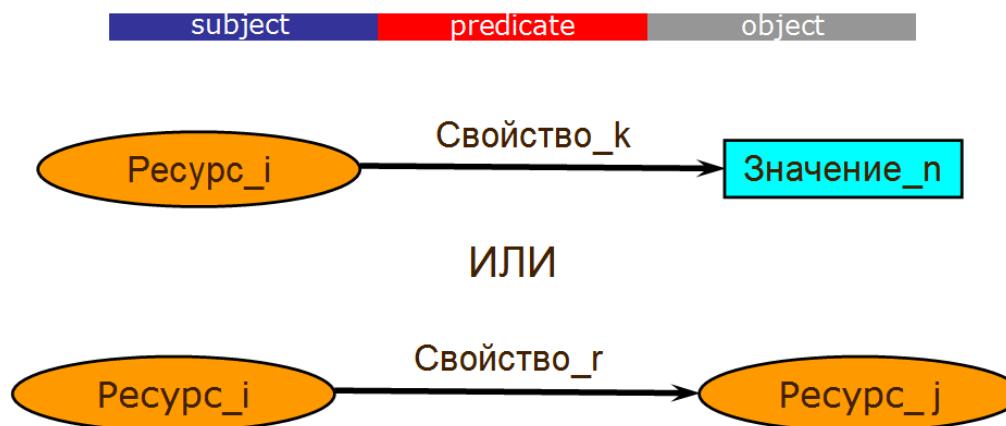
```
<country name="Ukraine">
    <president>Victor Yanukovich</president>
</country>
```

```
<president name="Victor Yanukovich">
    <country>Ukraine</country>
</president>
```

Обратная вложенность, но эквивалентный смысл!

Базис RDF

- Базовый элемент – триплет:



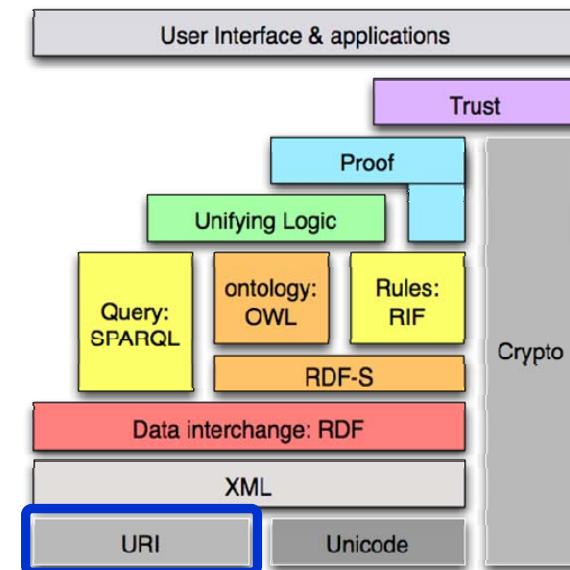
- Называется - **утверждение**
- Предложение о Yanukovich – такое утверждение
- RDF имеет сериализацию (синтаксис записанный) в XML
 - Этот синтаксис наследует преимущества XML
 - Существуют и ДРУГИЕ варианты представления синтаксиса RDF
 - ПР: ???

Базис RDF (2)

- Фундаментальные элементы концепции RDF:
 - Ресурсы
 - Свойства
 - Утверждения

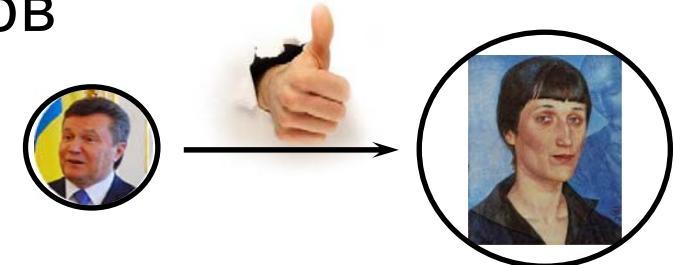
Фундамент RDF: Ресурс

- (Предыдущая тема)
- Ресурс – любой объект ...
 - Пр.: автор, книга, страна, президент, лектор, ...
- ... имеющий URI



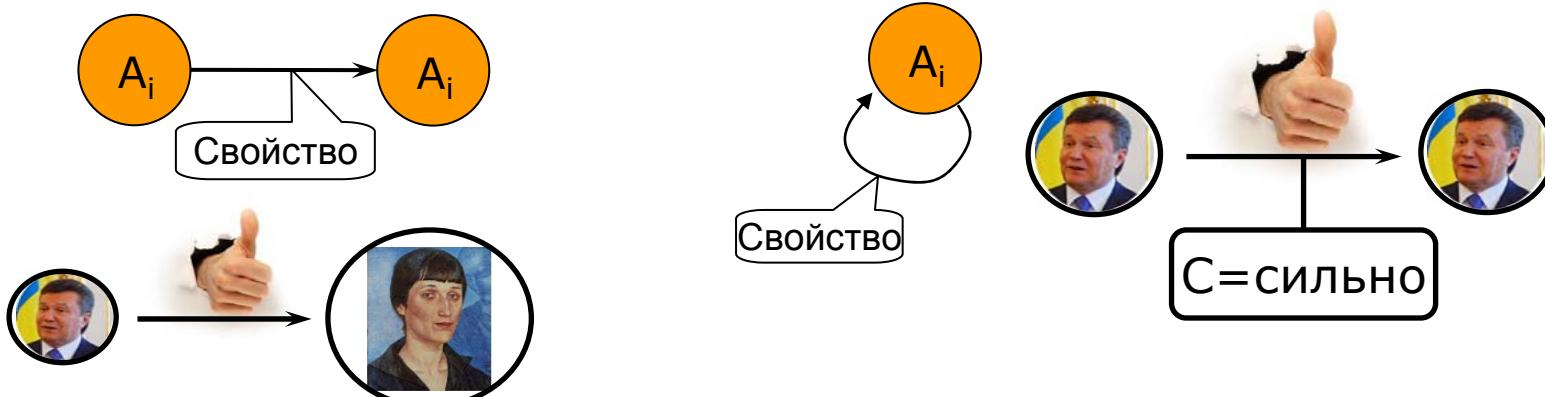
Фундамент RDF: Свойство

- Свойства – особый тип ресурсов
 - Описывают отношения между ресурсами
- Как и другие типы ресурсов, свойства идентифицируются при помощи URI
- Преимущества использования URI:
 - Глобальная и уникальная схема именования (идентификации)
 - Пр.: ==> уменьшает негативное влияние проблемы омонимов в распределенных представлениях данных

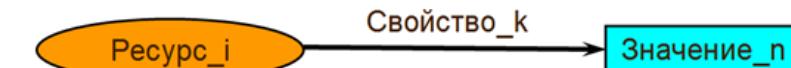


Фундамент RDF: Утверждение

- Утверждения специфицируют свойства ресурсов



- Формально, утверждение – триплет (ресурс-свойство-значение)



ИЛИ



- Значения могут быть ресурсами (URI) или литералами
- Литералы – атомарные значения (строки)

Три взгляда на утверждение

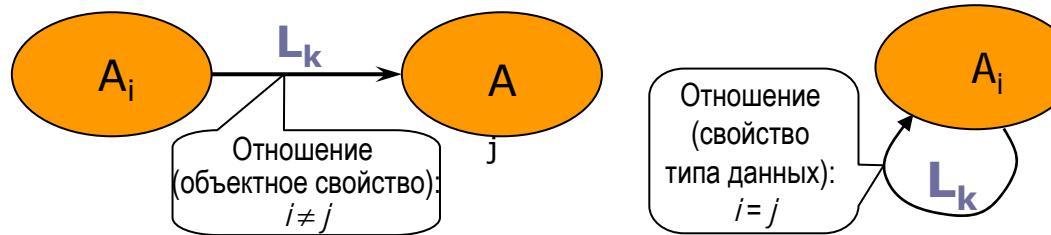
- Триплет
- Фрагмент графа
- Фрагмент XML кода

Т.о., документ RDF может рассматриваться как:

- Набор триплетов
- Граф (семантическая сеть)
- Документ XML

Утверждения как Триплеты

$$P(A_i, L_k, A_j) = \begin{cases} 1, & \text{если } A_i \text{ соединен отношением } L_k \text{ с } A_j; \\ 0, & \text{иначе} \end{cases}$$



(<http://www.cit.gu.edu.au/~db>,
<http://www.mydomain.org/site-owner>,
#David Billington)

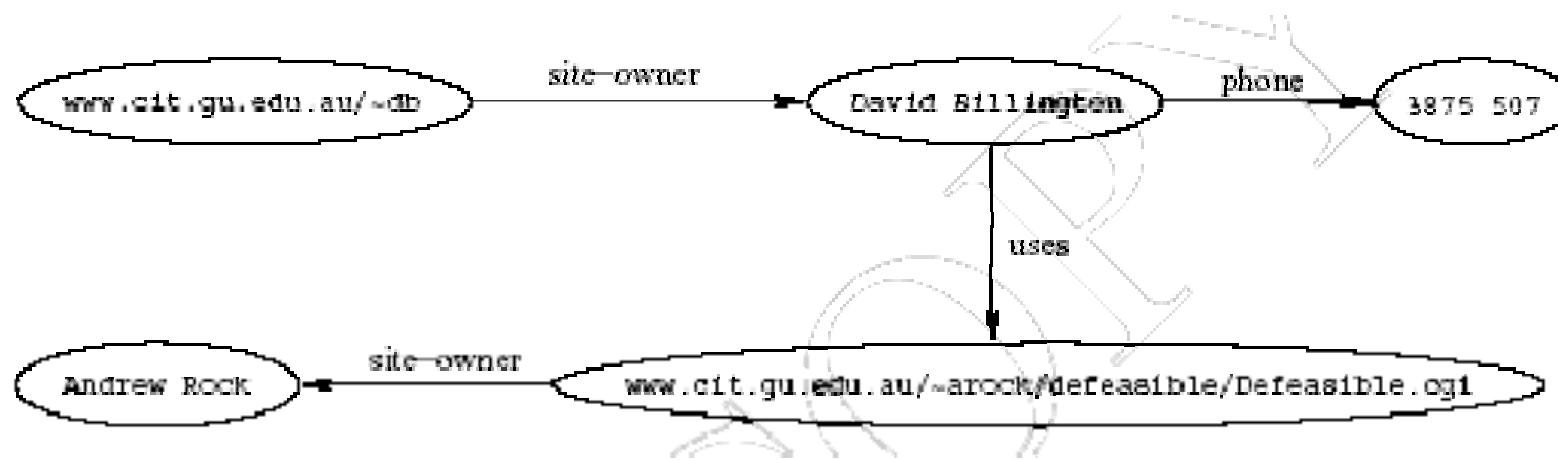
- Триплет $(A_i, P, A_j) \leftrightarrow$ логическая формула (предикат) $P(A_i, A_j)$
 - Бинарный предикат P устанавливает соответствие между объектами A_i, A_j
 - В RDF возможны только **бинарные предикаты** (свойства)

Утверждения как Графы



- Ориентированный помеченный граф
 - Вершины и дуги имеют метки
 - Направление:
 - от субъекта (ресурса)
 - к объекту (значению: ресурсу или литералу)
- В ИИ известны как **семантические сети**
- Значение в утверждении, если ресурс, может быть связано с другими ресурсами
 - Продолжение семантической сети

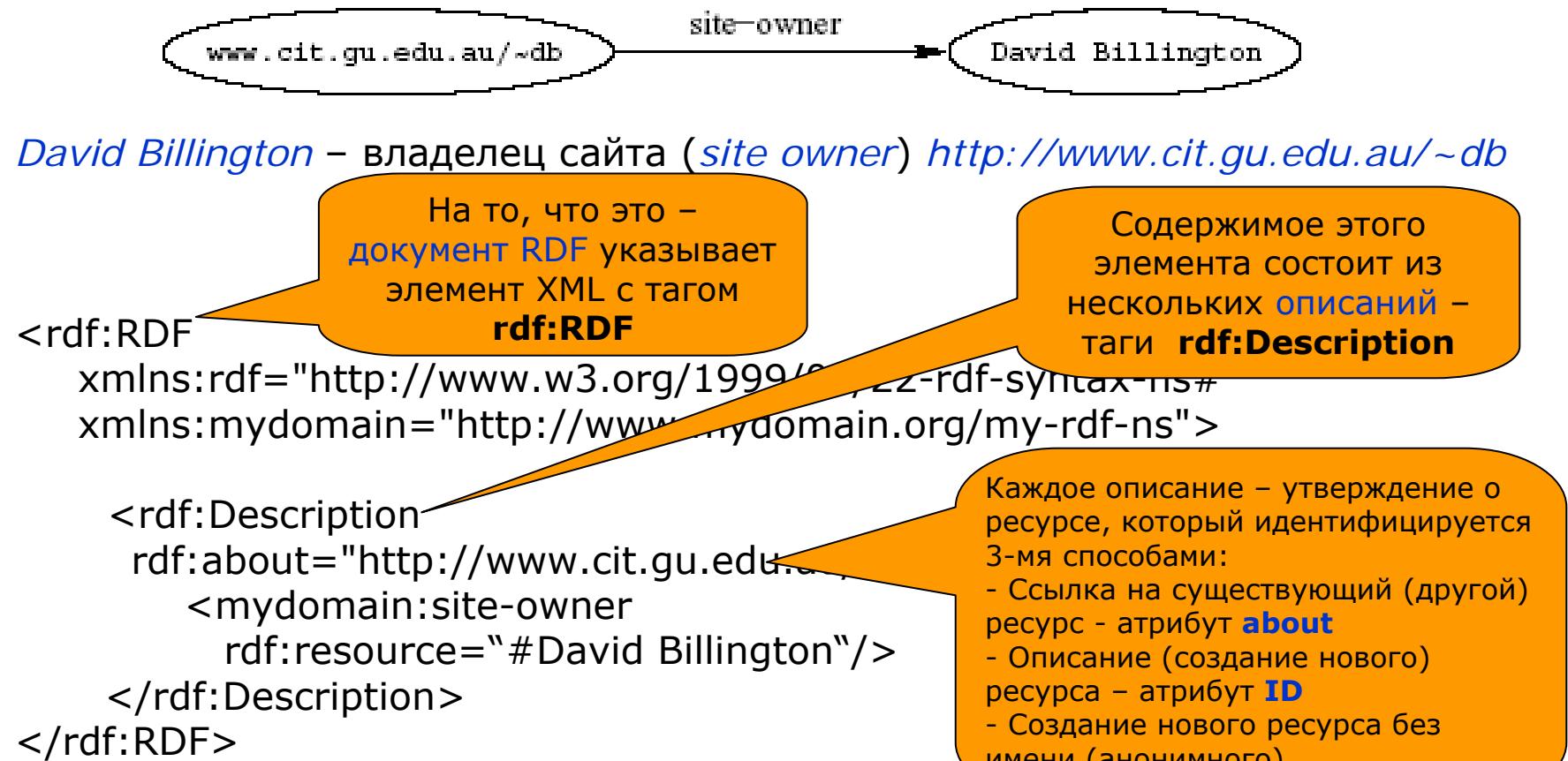
Набор Триплетов – Сем. Сеть



Утверждения как XML Код

- Графы – удобная и мощная нотация для понимания человеком
- Однако:
 - Семантический Веб требует представлений семантики, которые могут быть получены и обработаны ЭВМ
- Для этого используется представление RDF утверждений в виде кода на XML
 - Нужно помнить:
 - XML не является частью модели данных RDF
 - Пр.: сериализация (диалект) XML не имеет отношения к RDF

Утверждения RDF в XML



Утверждения RDF в XML

```
<rdf:Description  
    rdf:about="http://www.cit.gu.edu.au/~db">  
    <mydomain:site-owner  
        rdf:resource="#David Billington"/>  
</rdf:Description>
```

- Элемент **rdf:Description** содержит утверждение о ресурсе <http://www.cit.gu.edu.au/~db>
- Внутри элемента **rdf:Description**
 - Свойство специфицируется тагом
 - `mydomain:site-owner`
 - Содержимое тага представляет значение свойства
 - **`rdf:resource="#David Billington"`**

Материализация

Реификация

- В RDF возможно записывать утверждения об утверждениях
 - Пр.: **Grigoris** считает (**believes**), что **David Billington** является создателем (**creator**)
<http://www.cit.gu.edu.au/~db>
- Такие утверждения могут понадобиться для описания **отношения** (доверие, полагание) к другим утверждениям
- Техническое решение:
 - Назначить **уникальный идентификатор** утверждению
 - Такой идентификатор может в дальнейшем использоваться для ссылок на идентифицируемое утверждение

Материализация

Реификация

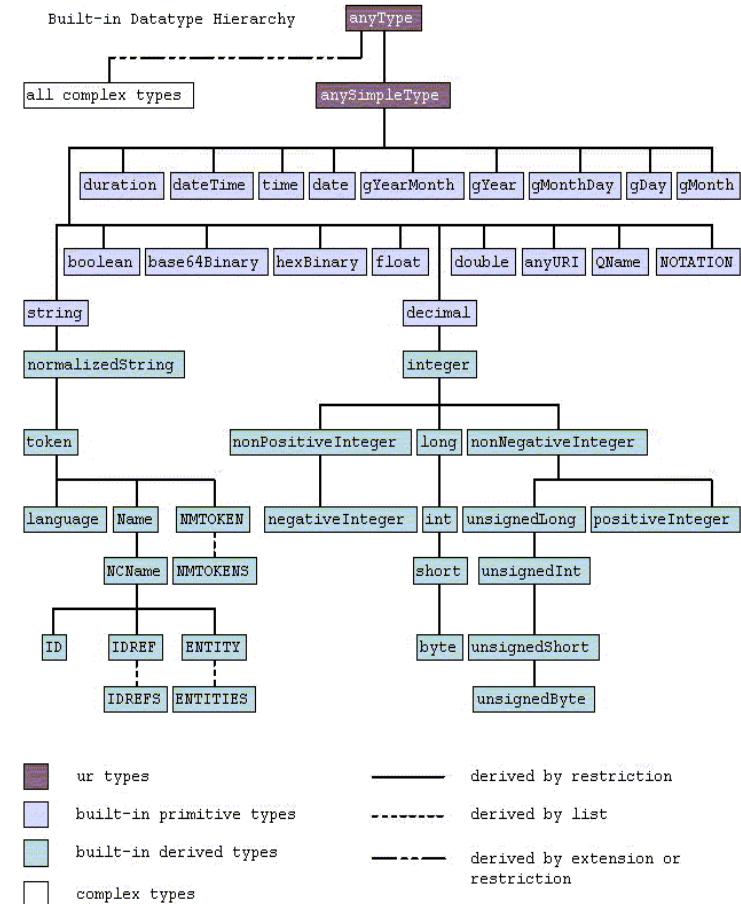
- Введем вспомогательный объект (утверждение - пр. **belief1**)
- Соотнесем его с каждой из трех частей оригинального утверждения через свойства **subject**, **predicate** и **object**
- В предыдущем примере оригинальное утверждение:
 - **Grigoris** считает (**believes**), что **David Billington** является создателем (**creator**)
<http://www.cit.gu.edu.au/~db>
 - subject **belief1**: David Billington
 - predicate **belief1**: creator
 - object of **belief1**: <http://www.cit.gu.edu.au/~db>

Типы Данных

- В RDF используются типизированные литералы

(#David Billington,
<http://www.mydomain.org/age>,
“27”^{http://www.w3.org/2001/XMLSchema#integer})

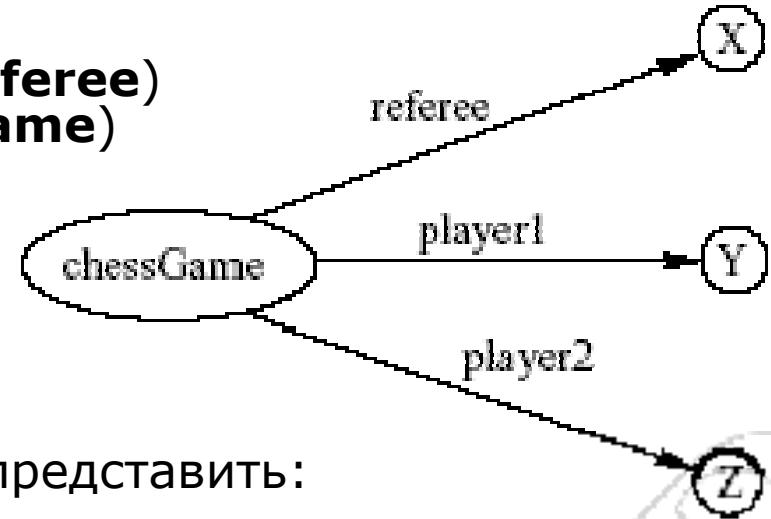
- ^-указывает (URI) тип литерала
- На практике наиболее часто используют типы данных, определенные в XML Schema
 - Разрешено использование **любой внешней** схемы, специфицирующей типы данных
- XML Schema определяет широкий спектр типов данных
 - Пр.: boolean, integer, float-point, time, date, и т.д.
 - <http://www.w3.org/TR/xmlschema-2/>



Критика RDF:

Бинарность Утверждений

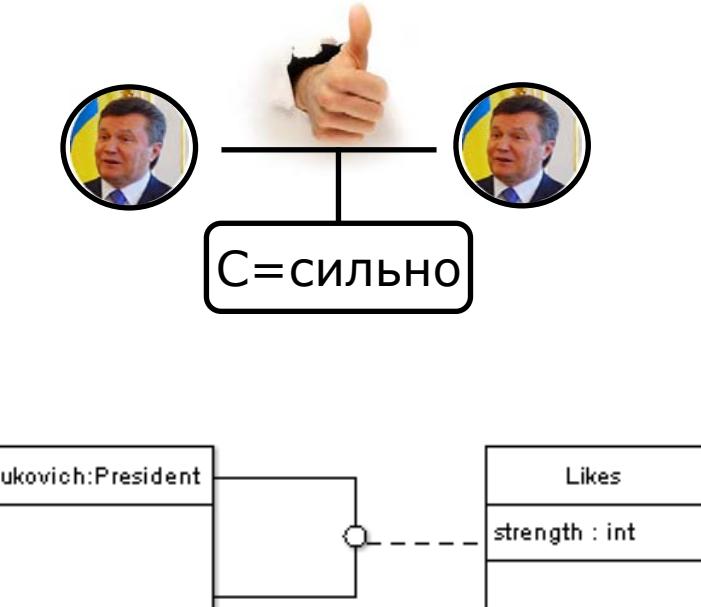
- RDF разрешает использовать **только бинарные** свойства
 - Это – серьезное ограничение – часто требуются свойства большей арности
 - Однако, такие свойства можно записывать при помощи бинарных
- Пр.: **referee(X,Y,Z)**
 - **Предложение:** X – судья (**referee**) в шахматной партии (**chessGame**) между игроками Y и Z
 - Введем:
 - Новый вспомогательный ресурс **chessGame**
 - Бинарные предикаты **referee**, **player1** и **player2**
 - Тогда **referee(X,Y,Z)** можно представить:



Критика RDF:

Свойства

- Свойства в RDF могут быть представлены как ресурсы
 - Имеют **URI**
 - Находятся в позиции **subject** RDF утверждения
 - Специфицируются **независимо** от ресурсов
 - Аналогия – отношения в реляционной модели данных
- Такая возможность дает **гибкость и унификацию**
- **ОДНАКО**, такой способ представления:
 - Необычен для:
 - Языков концептуального моделирования
 - Языков ОО программирования
 - Может осложнять работу при моделировании



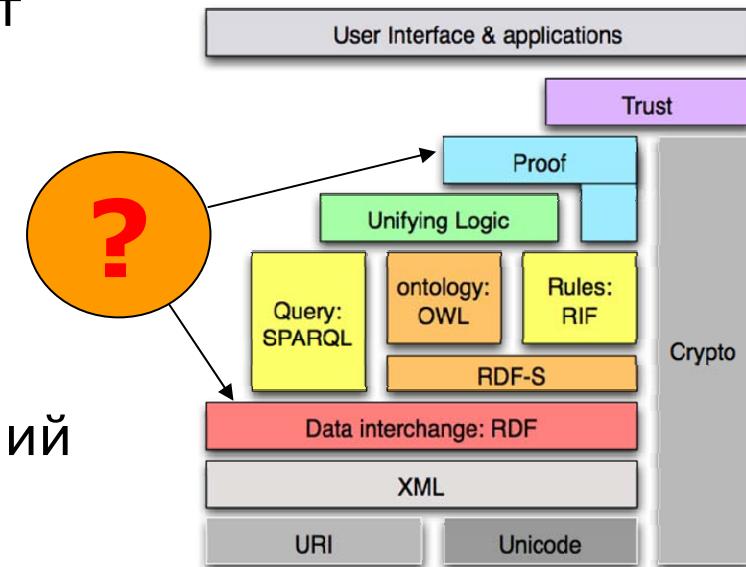
В XML синтаксисе ???

Критика RDF :

Материализация (Реификация)

Grigoris считает (**believes**), что **David Billington** является создателем (**creator**)
<http://www.cit.gu.edu.au/~db>

- Механизм реификации придает большую выразительную силу утверждениям
- **ОДНАКО**: Нужна ли такая выразительная сила в простом языке описания семантики Веб ресурсов?
- Спецификация утверждений об утверждениях потребуются на более высоких уровнях архитектуры Семантического Веба



Критика RDF: Резюме

- RDF имеет отличительные особенности
 - Которые можно критиковать
 - Или – использовать в описании Веб ресурсов
 - **Дело вкуса**
- **ОДНАКО** RDF :
 - Уже **де-факто стандарт**: <http://www.w3.org/RDF/>
 - «RDF is a standard model for data interchange on the Web»
 - Имеет **достаточную** выразительную силу
 - По крайней мере для более высоких уровней SW
- Использование RDF позволяет**недвусмысленно** отображать информацию в семантическую модель

Синтаксис RDF на базе XML

- Очень быстро ...
- Подробности на <http://www.w3.org/RDF/>
 - [RDF/XML Syntax Specification \(Revised\)](#)
Dave Beckett, ed.
 - [RDF Vocabulary Description Language 1.0: RDF Schema](#)
Dan Brickley, R.V. Guha, eds.
 - [RDF Primer](#)
Frank Manola, Eric Miller, eds.
 - [Resource Description Framework \(RDF\): Concepts and Abstract Syntax](#)
Graham Klyne, Jeremy Carroll, eds.
 - [RDF Semantics](#)
Patrick Hayes, ed.
 - [RDF Test Cases](#)
Jan Grant, Dave Beckett, eds.

XML-Синтаксис RDF

Структура Документа RDF

- Документ RDF специфицируется элементом **rdf:RDF**
 - Содержимое этого элемента – набор описаний **rdf:Description**
 - Пространства имен XML **xmlns: ...** используются для:
 - Обеспечения непротиворечивости (одинаковые имена – омонимы – в разных пространствах имен)
 - Пространства имен – документы RDF, определяющие ресурсы, которые могут использоваться повторно
 - Повторное использование позволяет строить большие распределенные коллекции представлений знаний

```
<rdf:RDF  
    xmlns:rdf="http://.../#"  
    xmlns:mydomain="...">  
  
    <rdf:Description ...>  
        ...  
    </rdf:Description>  
    ...  
</rdf:RDF>
```

XML-Синтаксис RDF

Пример

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:uni="http://www.mydomain.org/uni-ns">
    <rdf:Description rdf:about="949318">
        <uni:name>David Billington</uni:name>
        <uni:title>Associate Professor</uni:title>
        <uni:age rdf:datatype="&xsd:integer">27</uni:age>
    </rdf:Description>
    <rdf:Description rdf:about="CIT1111">
        <uni:courseName>Discrete Maths</uni:courseName>
        <uni:isTaughtBy>David Billington</uni:isTaughtBy>
    </rdf:Description>
    <rdf:Description rdf:about="CIT2112">
        <uni:courseName>Programming III</uni:courseName>
        <uni:isTaughtBy>Michael Maher</uni:isTaughtBy>
    </rdf:Description>
</rdf:RDF>
```

rdf:about или **rdf:ID**

- Элемент **rdf:Description** допускает
 - Атрибут **rdf:about** - указывает, что ресурс определен вне данного RDF документа
 - Атрибут **rdf:ID** - указывает, что ресурс определен данным описанием
- Формально, с точки зрения смысла утверждения, НЕТ разницы в месте определения
- Практическая польза:
 - Структуризация - определить:
 - Основное (общее) описание в одном месте
 - Добавочные свойства (акценты интерпретации, контекстуализация) – в других местах
- Повторное использование описаний

Элементы Описания Свойств

- Содержимое элементов **rdf:Description**

```
<rdf:Description rdf:about="CIT3116">
  <uni:courseName>
    Knowledge Representation
  </uni:courseName>
  <uni:isTaughtBy>
    Grigoris Antoniou
  </uni:isTaughtBy>
</rdf:Description>
```

???



- uni:courseName** и **uni:isTaughtBy** специфицируют две пары свойство – значение для одного ресурса (субъекта) **CIT3116**

- Два утверждения RDF
- Читать конъюнктивно

Типы Данных

- Атрибут **rdf:datatype="&xsd:integer"** используется для указания типа данных значения свойства возраст (age)

```
<rdf:Description rdf:about="949318">
    <uni:name>David Billington</uni:name>
    <uni:title>Associate Professor</uni:title>
    <uni:age rdf:datatype="&xsd:integer">27</uni:age>
</rdf:Description>
```

Типы Данных

```
<rdf:Description rdf:about="949318">
    <uni:name>David Billington</uni:name>
    <uni:title>Associate Professor</uni:title>
    <uni:age rdf:datatype="&xsd:integer">27</uni:age>
</rdf:Description>
```

- Свойство **age** имеет диапазон допустимых значений (*range*) "**&xsd:integer**"
 - Требуется также указывать тип этого свойства каждый раз при его использовании
 - Это требование обеспечивает, что обработчик RDF назначает корректный тип значению свойства, даже если он не «видел» соответствующее определение в RDF схеме
 - Типичный сценарий для обработки неструктурированных данных на WWW

Атрибут rdf:resource

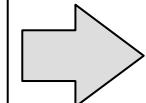
```
<rdf:Description rdf:about="949318">
    <uni:name>David Billington</uni:name>
    <uni:title>Associate Professor</uni:title>
    <uni:age rdf:datatype="&xsd:integer">27<uni:age>
</rdf:Description>
<rdf:Description rdf:about="CIT1111">
    <uni:courseName>Discrete Maths</uni:courseName>
    <uni:isTaughtBy>David Billington</uni:isTaughtBy>
</rdf:Description>
```

} лектор

} курс

- Связь между курсом и лектором не специфицирована формально
- Однако, существует – неявно – использовано одно имя (значение свойства)
- Однаковое имя – возможно, совпадение
- Для **ЯВНОГО** указания, что две сущности – одно и то же используют атрибут **rdf:resource**

```
<rdf:Description  
    rdf:about="949318">  
    <uni:name>  
        David Billington  
    </uni:name>  
    <uni:title>  
        Associate Professor  
    </uni:title>  
</rdf:Description>  
<rdf:Description  
    rdf:about="CIT1111">  
    <uni:courseName>  
        Discrete Mathematics  
    </uni:courseName>  
    <uni:isTaughtBy>  
        David Billington  
    </uni:isTaughtBy>  
</rdf:Description>
```



```
<rdf:Description  
    rdf:about="CIT1111">  
    <uni:courseName>  
        Discrete Mathematics  
    </uni:courseName>  
    <uni:isTaughtBy>  
        rdf:resource="949318"/>  
</rdf:Description>  
<rdf:Description  
    rdf:about="949318">  
    <uni:name>  
        David Billington  
    </uni:name>  
    <uni:title>  
        Associate Professor  
    </uni:title>  
</rdf:Description>
```

Указание на Ресурсы,

Специфицированные Извне

- Пр., как указать, где специфицирован ресурс CIT1111?
 - **http://www.mydomain.org/uni-ns#CIT1111**
 - В значении атрибута **rdf:about**
 - <rdf:Description rdf:about="http://www.mydomain.org/uni-ns#CIT1111">
 - **www.mydomain.org/uni-ns** - URI, указывающий ресурс (документ RDF), в котором находится спецификация CIT1111
 - Эта спецификация использует атрибут **rdf:ID**, значение которого используется для указаний в **rdf:about**
 - <rdf:Description rdf:ID="CIT1111">

Вложенные Описания

```
<rdf:Description rdf:about="CIT1111">
    <uni:courseName>Discrete Maths
    </uni:courseName>
    <uni:isTaughtBy>
        <rdf:Description rdf:ID="949318">
            <uni:name>David Billington</uni:name>
            <uni:title>Associate Professor</uni:title>
        </rdf:Description>
    </uni:isTaughtBy>
</rdf:Description>
```

- Описания (**rdf:Description**) могут специфицироваться внутри других описаний
- Важно: другие курсы, такие как **CIT3112**, могут также ссылаться на локально специфицируемый ресурс с **rdf:ID="949318"**
 - Хотя описание может быть определено внутри другого описания, его сфера действия (видимость) глобальна

Элемент rdf:type

Структурирование Документов RDF

```
<rdf:Description rdf:ID="CIT1111">
    <rdf:type
        rdf:resource="http://www.mydomain.org/uni-
        ns#course"/>
        <uni:courseName>Discrete Maths</uni:courseName>
        <uni:isTaughtBy rdf:resource="#949318"/>
</rdf:Description>
<rdf:Description rdf:ID="949318">
    <rdf:type
        rdf:resource="http://www.mydomain.org/uni-
        ns#lecturer"/>
        <uni:name>David Billington</uni:name>
        <uni:title>Associate Professor</uni:title>
</rdf:Description>
```

Ресурс
rdf:ID="CIT1111"
имеет тип "course".
Тип определен извне.

Ресурс
rdf:ID="949318"
имеет тип "lecturer".
Тип определен извне.

Не путать с типом данных!!!

Сокращенный Синтаксис

- **Правила сокращения (упрощения):**
 1. Элементы, описывающие свойства и не имеющие вложенных в них элементов (детей), могут заменяться атрибутами XML
 2. Для элементов – описаний (**rdf:Description**), содержащих указание типа (**rdf:type**), можно использовать имя, указанное в **rdf:type** вместо **rdf:Description**
- Использование этих правил порождает **синтаксические вариации** одного и того же утверждения RDF
 - Эти вариации эквивалентны с точки зрения модели данных RDF, хотя и имеют отличающийся синтаксис в XML сериализации

Сокращенный Синтаксис:

Пример

```
<rdf:Description rdf:ID="CIT1111">
    <rdf:type
        rdf:resource="http://www.mydomain.org/uni-
        ns#course"/>
    <uni:courseName>Discrete Maths</uni:courseName>
    <uni:isTaughtBy rdf:resource="#949318"/>
</rdf:Description>
```

Сокращенный Синтаксис:

Пример - Применяем Правило 1

```
<rdf:Description rdf:ID="CIT1111">
    <rdf:type
        rdf:resource="http://www.mydomain.org/
        uni-          ns#course"/>
        <uni:courseName>Discrete
        Maths</uni:courseName>
        <uni:isTaughtBy
            rdf:resource="#949318"/>
</rdf:Description>
```

Сокращенный Синтаксис:

Пример - Применяем Правило 2

```
<rdf:Description rdf:ID="CIT1111"  
    uni:courseName="Discrete Maths">  
    <rdf:type rdf:resource="http://www.mydomain.org/uni-  
        ns#course"/>  
    <uni:isTaughtBy rdf:resource="#949318"/>  
</rdf:Description>
```

2. Для элементов – описаний (**rdf:Description**), содержащих указание типа (**rdf:type**), можно использовать имя, указанное в **rdf:type** вместо **rdf:Description**

```
<uni:course rdf:ID="CIT1111"  
    uni:courseName="Discrete Maths">  
    <uni:isTaughtBy rdf:resource="#949318"/>  
</uni:course>
```

Контейнеры

- Собирают вместе несколько ресурсов или атрибутов о которых мы хотим специфицировать утверждение как о целом
- Пр.: мы хотим записать утверждение о курсах, читаемых одним лектором
 - Аналогия: связь один-ко-многим в РМД
- Содержимое элементов контейнера именуется **rdf:_1**, **rdf:_2**, и т.д.
 - Альтернативно можно использовать **rdf:li**

Три Типа Элементов Контейнера

- **rdf:Bag** – неупорядоченный контейнер, разрешающий несколько вхождений
 - Пр.: документы/файлы в папке
- **rdf:Seq** – упорядоченный контейнер, an ordered container, разрешающий несколько вхождений, порядок следования которых специфицирован
 - Пр.: модули в рамках курса, алфавитный список студентов группы
- **rdf:Alt** – набор альтернатив
 - Пр.: основной URL видео файла и его зеркала, переводы документа на различніе языки

Пример с rdf:Bag

```
<uni:lecturer  
    rdf:ID="949352" uni:name="Grigoris  
Antoniou" uni:title="Professor">  
    <uni:coursesTaught>  
        <rdf:Bag>  
            <rdf:_1  
rdf:resource="#CIT1112"/>  
            <rdf:_2  
rdf:resource="#CIT3116"/>  
        </rdf:Bag>  
    </uni:coursesTaught>
```

Пример с rdf:Alt

```
<uni:course rdf:ID="CIT1111"  
            uni:courseName="Discrete Mathematics">  
    <uni:lecturer>  
        <rdf:Alt>  
            <rdf:li rdf:resource="#949352"/>  
            <rdf:li rdf:resource="#949318"/>  
        </rdf:Alt>  
    </uni:lecturer>  
</uni:course>
```

Атрибут `rdt:ID` для Контейнера

```
<uni:lecturer rdf:ID="949318"  
              uni:name="David Billington">  
  <uni:coursesTaught>  
    <rdf:Bag rdf:ID="DBcourses">  
      <rdf:_1 rdf:resource="#CIT1111"/>  
      <rdf:_2 rdf:resource="#CIT3112"/>  
    </rdf:Bag>  
  </uni:coursesTaught>  
</uni:lecturer>
```

Коллекции в RDF

- Ограничение контейнера заключается в невозможности «закрыть» его как множество – т.е. записать:
 - «Это **ВСЕ** элементы в данном контейнере»
- RDF позволяет специфицировать наборы, содержащие только специфицированные элементы – как **RDF коллекции (collections)**
- **RDF коллекция**
 - **Списочная структура** в RDF графе
 - Специфицируется с помощью следующих элементов (словаря):
 - **rdf>List**, **rdf:first**, **rdf:rest** и **rdf:nil**

Коллекции в RDF

Сокращенный Синтаксис

- **Значение "Collection" атрибута `rdf:parseType`:**

```
<rdf:Description rdf:about="#CIT2112">
    <uni:isTaughtBy rdf:parseType="Collection">
        <rdf:Description rdf:about="#949111"/>
        <rdf:Description rdf:about="#949352"/>
        <rdf:Description rdf:about="#949318"/>
    </uni:isTaughtBy>
</rdf:Description>
```

Реификация

- Утверждения о (других) утверждениях
- Необходим идентификатор утверждения – для ссылки
- RDF разрешает подобные ссылки через механизм реификации, который трансформирует утверждение в ресурс

Реификация

Пример

```
<rdf:Description rdf:about="#949352">  
    <uni:name>Grigoris Antoniou</uni:name>  
</rdf:Description>
```

- **Реифицируется** как:

```
<rdf:Statement rdf:ID="StatementAbout949352">  
    <rdf:subject rdf:resource="#949352"/>  
    <rdf:predicate rdf:resource="http://www.mydomain.org/  
        uni-ns#name"/>  
    <rdf:object>Grigoris Antoniou</rdf:object>  
</rdf:Statement>
```

Реификация

- **rdf:subject**, **rdf:predicate** и **rdf:object** позволяют получить доступ к частям утверждения
- **ID** утверждения может быть использован для ссылок на утверждение – аналогично любому другому описанию
- Мы пишем **rdf:Description**, если в дальнейшем **НЕ ХОТИМ** ссылаться на (строить утверждения о) данное утверждение
- Мы пишем **rdf:Statement**, если в дальнейшем **ХОТИМ** ссылаться на (строить утверждения о) данное утверждение

object

predicate

subject

Вопросы



Следующая лекция

в этой теме

Т3: Семантика ресурсов: RDF Schema, RDFa

Subject predicate object



Acknowledgements

- This **teaching material** is based on:
 - **Grigoris Antoniou's and Frank van Harmelen's** support material for Chapter 3 of their **Semantic Web Primer**
 - <http://www.ics.forth.gr/isl/swprimer/presentations/Chapter3.ppt>
- Some slides were **adopted** or **adapted** from various presentations by (alphabetically):
 - Tim Berners-Lee, Steve Bratt, Vadim Ermolayev, Tim Finin